

REMARKS/ARGUMENTS

I. Introduction:

Claims 1-38 are currently pending.

II. Claim Rejections under 35 U.S.C. 103:

Claims 1, 6-8, 10, 12-15, 18, 21-29, 31-34, and 36-38 stand rejected under 35 U.S.C. 103(a) as being unpatentable over U.S. Patent No. 6,351,775 (Yu) in view of U.S. Patent No. 6,249,836 (Downs et al.).

Yu is directed to load balancing across servers and uses a partitioning method to group object identifiers into classes. Clients are required to maintain a class-to-server assignment table to map each class into a server selection. The class-to-server assignment tables can change dynamically at any time as the workload varies. An important aspect of the Yu invention is that the system not only balances a load across a cluster of servers, but also optimizes a cache hit ratio in a given server by localizing identical object requests. Routing requests for the same object to a single server node results in a better cache hit probability at the same server node. In order to optimize cache hit, object identifiers are grouped into classes which are identified in the class-to-server assignment table to map each class into a server selection. The class-to-server assignment table assigns each class to a virtual server and routers dynamically map each virtual server to one of the real servers in the cluster.

Downs et al. disclose a method and apparatus for providing remote processing of a task over a network. A resource provider first communicates to the owner of a server an intent to sell its resources. The client that needs processing for a particular task then sends a request for service to the server. The server matches the request for service to one of the resource providers and transfers the task to that resource provider.

The resource provider either performs the task to completion or notifies the server that it cannot complete the task. If the task is not completed the client can be directed to another resource provider, in which case the connection is terminated before completion of the request.

Applicants respectfully submit that that claim 1 is patentable over Yu and Downs et al., which do not disclose providing a persistent connection between a client and a server, binding a primary virtual server to a set of real servers, or receiving a request from a client for connection to the primary virtual server, as set forth in claim 1.

Applicants' invention, as set forth in claim 1, provides binding between a primary virtual server and a set of real servers. In contrast, Yu groups map object identifiers into classes and assigns each class to a virtual server. A router then dynamically maps each virtual server to one of the actual servers in a cluster. As noted at col. 11, lines 61-62, the number of virtual servers is greater than the actual number of servers in the server cluster.

Since each client in the load balancing system of Yu maintains a class-to-server assignment table, there is no need to send a request for connection to a primary virtual server. The load balancing is performed by updating the assignment table (col. 6, lines 30-33). The specific server to which an object request is sent is determined from the assignment table. Thus, Yu teaches away from sending a request to a primary virtual server. Similarly, since the server information is contained within the assignment table, there is no need to send a redirect message to the client specifying a selected real server.

Furthermore, Yu monitors the load of each server and dynamically updates the class-to-server assignment to improve load balancing. These updates can be provided at any time, including while a client and server are conducting a session or transaction. (See, for example, discussion of reassignment routine at col. 9, lines 63-65, and col. 10, lines 4-42, and col. 12, lines 36-48 "if a server receives a request from a requester that is

no longer assigned to that server, the server will inform the requester of the server to which future requests should be issued.”). The system of Yu can update assignment tables every minute to assign the requester to a less loaded server (col. 12, lines 49-63). Thus, a requested node can be connected to a number of different servers during a transaction or session.

The Examiner cites Downs et al. as showing connection of a server and a client for the duration of a transaction. The Downs et al. system provides remote processing of a task over a network. Once a task is assigned to a resource provider, the task may be terminated before completion if it does not have adequate resources to complete the task. If the task is not completed, the client can be directed to another resource provider, in which case the connection is terminated before completion of the request.

Applicants’ invention is particularly advantageous in that it provides a persistent (or “sticky”) connection between a user and a server. The sticky connection allows a controller or load balancer to direct each client connection in a session to the same server so that all requests from a given client are redirected to the same server and the client remains attached to a single server for the duration of the session between the client and the server. Since HTTP, for example, does not carry state information for applications such as shopping baskets, financial transactions, or interactive games, it is important for the user to be mapped to the same server for each request until a transaction or session is complete.

Accordingly, claim 1 is submitted as patentable over the prior art of record. Claims 2-12 and claims 24-31, depending either directly or indirectly from claim 1, are submitted as patentable for the same reasons as claim 1.

Claims 7 and 12 are further submitted as patentable over Yu which does not disclose sending an HTTP redirect message to the client specifying the selected real server, or receiving a request at a local director, as set forth in claims 7 and 12,

respectively. As discussed above, each requester node of Yu maintains a server assignment table to map each class into a server selection. There is, therefore, no need to send a request to a local director to select a server or send a redirect message to the client to inform the client of the selected server.

Claims 13, 14, 18, 21, 22, and 23 specify that the connection between the client and server is a persistent connection and are submitted as patentable for the reasons discussed above with respect to claim 1.

Claims 32-33, 15-17, 19-20 and 37-38, and 34-36, depending either directly or indirectly from claims 13, 14, 18, and 21, respectively, are submitted as patentable for the same reasons as their respective independent claims.

Claims 2-5, 9, 16-17, 19-20, and 35 stand rejected under 35 U.S.C. 103(a) as being unpatentable over Yu in view of Downs et al. and further in view of U.S. Patent No. 6,609,213 (Nguyen et al.). The Nguyen et al. patent was filed on August 10, 2000. The present patent application was filed on June 30, 2000. Therefore, the Nguyen et al. patent is not prior art with regard to the present application. Applicants request that all rejections based on the Nguyen et al. reference be withdrawn.

Appl. No. 09/608,789
Amd. Dated March 10, 2005
Reply to Final Office Action of December 22, 2004

III. Conclusion:

For the foregoing reasons, Applicants believe that all of the pending claims are in condition for allowance and should be passed to issue. If the Examiner feels that a telephone conference would in any way expedite prosecution of the application, please do not hesitate to call the undersigned at (408) 446-8695.

Respectfully submitted,



Cindy S. Kaplan
Reg. No. 40,043

RITTER, LANG & KAPLAN LLP
12930 Saratoga Ave., Suite D1
Saratoga, CA 95070
Tel: 408-446-8690
Fax: 408-446-8691